

COLLOQUE NATIONAL SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 26 au 30 AVRIL 1977

ESTIMATION LINÉAIRE ADAPTATIVE - APPLICATION AUX TRANSMISSIONS DE DONNÉES

ODILE MACCHI

Laboratoire des Signaux et Systèmes - CNRS-ESE - Plateau du Moulon 91190 GIF/YVETTE (France)

RESUME

Le problème de l'estimation optimale d'un message perturbé par un bruit -non nécessairement additif- trouve sa solution par la résolution d'une équation de Wiener-Hopf. Même discrétisée, cette équation n'est pas facile à résoudre. En outre elle fait intervenir des moments du signal observé qui sont souvent inconnus. Dans ce cas on peut utiliser des algorithmes d'apprentissage autoadaptatifs pour trouver itérativement l'estimateur optimal. Dans cet article, on précise les propriétés d'autoadaptativité et de convergence de deux algorithmes couramment utilisés. On généralise leur emploi au cas pratique intéressant de signaux et de messages complexes comme dans la modulation d'amplitude de deux porteuses en quadrature.

On étudie en détail le cas des transmissions de données. Dans ce cas on démontre que les conditions de convergence sont satisfaites lorsque le canal de transmission est un canal filtrant, affecté de gigue stationnaire et qui ajoute un bruit (la gigue est un bruit multiplicatif complexe). Le bruit est de nature assez complexe puisqu'il résulte à la fois du bruit additif, des interférences entre données adjacentes, et du bruit multiplicatif.

SUMMARY

Optimum estimation of a message imbedded in a -non necessarily additive- noise implies the resolution of a Wiener-Hopf equation, which is uneasy even in discrete time. Moreover some of the statistical averages involved may be unknown. If so, stochastic autoadaptive learning algorithms can be used which provide iteratively the optimum estimator.

In this paper, we specify the properties of adaptivity and convergence of two such algorithms, among the most often used. We extend the proof of their convergence to the important practical situation of complex signals and messages, as in quadrature amplitude modulation.

We detail the case of data transmissions. We show that the convergence conditions are then fulfilled for a filtering channel affected by stationary phase error ("phase jitter") and additive noise. In that transmission, the noise results from the complex multiplicative noise (jitter), the additive noise, and the intersymbol interferences.



ESTIMATION LINÉAIRE ADAPTATIVE - APPLICATION AUX TRANSMISSIONS DE DONNÉES

I - INTRODUCTION

On voit de plus en plus souvent intervenir dans les systèmes modernes de réception (radar, sonar, transmission de données, etc...) des algorithmes stochastiques destinés à adapter le récepteur au contexte de la transmission. Ces algorithmes doivent leur succès à leur grande simplicité de mise en oeuvre, basée sur la récursivité. Ils réalisent automatiquement l'adaptation du récepteur (auto-adaptivité). En particulier nombre de ces algorithmes sont destinés à fournir adaptativement un estimateur optimal. Mais le mode de convergence de ces algorithmes stochastiques n'est pas à l'heure actuelle clairement établi.

Dans cet article, nous considérons le problème classique de l'estimation d'un message émis $a(t)$, aléatoire, à partir d'un signal reçu observé $x(t)$, lui aussi aléatoire, et nous discutons le mode de convergence des algorithmes stochastiques actuellement utilisés dans la pratique pour construire simplement l'estimateur optimal.

Nous supposons que les signaux traités sont complexes afin de rendre compte d'un contexte fréquent en communications, particulièrement en transmission de données. C'est le cas lorsque le signal émis comporte deux porteuses en quadrature modulées respectivement par deux messages réels $a_1(t)$ et $a_2(t)$ (modulation MAQ). Le signal reçu comporte alors aussi deux composantes réelles en quadrature $x_1(t)$ et $x_2(t)$, et nous définirons l'observation complexe et le message complexe par

$$(1) \quad x(t) = x_1(t) + jx_2(t), \quad a(t) = a_1(t) + ja_2(t).$$

Enfin nous considérons un contexte entièrement numérique, où l'ensemble des signaux et des messages sont échantillonnés, fournissant la suite $\{\dots, x_j, \dots\}$ des observations et la suite $\{\dots, a_k, \dots\}$ des messages émis.

II - RAPPEL SUR LE FILTRAGE OPTIMAL

Si l'on dispose de $(L+1)$ observations passées x_{k-L}, \dots, x_k et de N observations futures x_{k+1}, \dots, x_{k+N} pour estimer a_k et si l'ensemble des signaux est stationnaire, un estimateur linéaire H est caractérisé par un filtrage numérique non récursif, invariant au cours du temps, portant sur le vecteur observation à l'instant k

$$(2) \quad \vec{X}_k^T = (x_{k-L}, \dots, x_k, \dots, x_{k+N}).$$

Si \vec{H} est le vecteur de filtrage -vecteur à $(N+L+1)$ coordonnées complexes - l'estimée de a_k vaut

$$(3) \quad y_k = \vec{X}_k^T \vec{H},$$

et

$$(4) \quad e_k(\vec{H}) = a_k - \vec{X}_k^T \vec{H}$$

est l'erreur d'estimation. On a l'habitude pour optimiser l'estimateur de minimiser l'erreur quadratique moyenne :

$$(5) \quad \epsilon(\vec{H}) = E(|e_k(\vec{H})|^2) \quad \text{minimum.}$$

En écrivant le gradient de $|e_k(\vec{H})|^2$ par rapport au vecteur \vec{H}_1 (resp. \vec{H}_2) des coordonnées réelles (resp. imaginaires) de \vec{H} , et en regroupant ces deux gradients en un seul, à coordonnées complexes, on obtient le gradient complexe de $|e_k(\vec{H})|^2$ par rapport à \vec{H} . On trouve

$$(6) \quad \nabla_{\vec{H}} [|e_k(\vec{H})|^2] = -2a_k \vec{X}_k^* + 2\vec{X}_k^* \vec{X}_k^T \vec{H}.$$

Tout vecteur \vec{H} solution de (5) est donc solution de l'équation

$$(7) \quad E(\vec{X}_k^* \vec{X}_k^T) \vec{H} = E(\vec{X}_k^* a_k).$$

C'est une équation de Wiener-Hopf discrétisée qui fait apparaître la matrice de covariance

$$(8) \quad \mathcal{R}_k = E(\vec{X}_k^* \vec{X}_k^T)$$

du signal observé (matrice que nous supposons inversible); et le vecteur d'intercorrélacion

$$(9) \quad \vec{T} = E(\vec{X}_k^* a_k)$$

entre l'observation et le message à estimer.

III - ALGORITHMES D'APPRENTISSAGE

L'optimisation directe de l'estimateur nécessite la résolution de l'équation (7). Dans un premier temps il faut évaluer les moments \mathcal{R}_k et \vec{T} a priori inconnus, parfois même lentement variables. Cette évaluation doit se faire à partir des vecteurs \vec{X}_k et, dans une période d'apprentissage, à partir d'une suite

ESTIMATION LINÉAIRE ADAPTATIVE - APPLICATION AUX TRANSMISSIONS DE DONNÉES

de messages tests a_k . Dans un deuxième temps il faut inverser la matrice R_k pour évaluer $R_k^{-1} \vec{T}$.

Tous les algorithmes d'apprentissage [1] - [3] usuels ont pour but de réaliser ces deux opérations globalement, et de manière itérative. Ils dérivent tous de l'algorithme du gradient

$$(10) \quad \vec{H}_{k+1} = \vec{H}_k - \mu_k \nabla_{\vec{H}} \epsilon(\vec{H}_k),$$

soit d'après (6)

$$(11) \quad \vec{H}_{k+1} = \vec{H}_k + \mu_k [-E(\vec{X}_k^* \vec{X}_k^T) \vec{H}_k + E(a_k \vec{X}_k^*)].$$

On sait [4] que cet algorithme converge vers le vecteur \vec{H}_* solution de (7) pour des constantes μ_k bien choisies.

Dans le cas où les moyennes R_k et \vec{T} sont inconnues, on ne peut utiliser (11), et l'on supprime tout simplement la moyenne d'ensemble dans le gradient ; d'où l'algorithme stochastique

$$(12) \quad \vec{H}_{k+1} = \vec{H}_k + \mu_k \vec{X}_k^* (-\vec{X}_k^T \vec{H}_k + a_k).$$

L'idée sous-jacente est que les itérations successives effectuent une moyenne temporelle qui, dans le cas ergodique, s'identifie à la moyenne d'ensemble et donc, qu'à la limite les algorithmes (12) et (11) sont les mêmes.

Dans le cas de signaux et messages réels, on peut encore simplifier l'algorithme (12) en utilisant seulement le signe du reste (voir [5]).

L'algorithme (12) utilise l'écart $e_k(\vec{H}_k)$ entre la vraie valeur a_k du message, et l'estimée $\vec{X}_k^T \vec{H}_k$ qu'en donne le filtre \vec{H}_k à sa $k^{\text{ième}}$ itération, pour adapter ce filtre par l'intermédiaire du reste $\vec{X}_k^* e_k$. Cette adaptation constitue un apprentissage du filtre qui minimise non pas $|e_k|^2$, grandeur aléatoire, mais sa valeur moyenne. C'est donc un apprentissage de \vec{H}_* .

Dans la pratique, l'algorithme (12) est utilisé sous deux formes, avec des coefficients μ_k décroissants ou constants :

$$(13) \quad \mu_k \geq 0, \quad \sum_k \mu_k = \infty, \quad \sum_k \mu_k^2 < \infty ;$$

$$(14) \quad \mu_k = \mu > 0.$$

Ces coefficients sont généralement appelés "pas" de l'algorithme ("step-size"). Nous allons maintenant expliciter la différence entre ces deux types d'algorithme en comparant leur qualité d'adaptation.

IV - COMPARAISON DE L'ADAPTATIVITE DES ALGORITHMES A PAS DECREOISSANTS ET A PAS CONSTANTS

Pour mieux expliciter la différence entre les deux algorithmes, supposons que la matrice R_k est connue, mais non le vecteur \vec{T} . L'algorithme d'apprentissage (12) devient donc

$$(15) \quad \vec{H}_{k+1} = \vec{H}_k + \mu_k (-R_k \vec{H}_k + \vec{X}_k^* a_k),$$

qui peut s'écrire

$$(16) \quad \vec{H}_{k+1} = D_{\nu}^k \{\mu_j\} \vec{H}_0 + \sum_{p=0}^k D_{\nu}^k \{\mu_j\} \vec{X}_p^* a_p$$

où les matrices $D_{\nu}^k \{\mu_j\}$ et $D_{\nu}^k \{\mu_j\}$ dépendent de la suite des coefficients μ_j . L'algorithme (15) réalise donc une sorte de filtrage sur la suite des vecteurs $a_p \vec{X}_p^*$. Posons par définition

$$(17) \quad \lambda = \sup \{ \text{valeurs propres de } R_k \}, \quad \nu = \inf \{ \text{valeurs propres de } R_k \}$$

Considérons d'abord l'algorithme (12)-(13) à pas décroissants, avec par exemple

$$(18) \quad \mu_j = \frac{\rho}{j+1}, \quad \rho > 0 \quad \nu < \rho^{-1} \leq \lambda.$$

On montre qu'alors on a les inégalités suivantes entre matrices définies positives :

$$(19) \quad \frac{\rho}{k+1} I \leq D_{\nu}^k \{\mu_j\} \leq \frac{\alpha \rho}{k \nu \rho} I$$

où I est la matrice identité, α_{ρ} est une constante positive.

L'inégalité de gauche montre que l'algorithme effectue une combinaison à coefficients d'égale importance des vecteurs successifs $\vec{X}_p^* a_p$, les observations les plus anciennes intervenant autant que les observations les plus récentes dans le calcul de l'estimateur. L'inégalité de droite montre que tous ces coefficients tendent conjointement mais lentement vers zéro lorsque k tend vers l'infini, en $k^{-\rho \nu}$.

Si les propriétés statistiques du vecteur $\vec{X}_k^* a_k$ changent, l'estimateur s'adapte aux nouvelles propriétés statistiques au bout d'un certain temps τ_d , que l'on appellera "temps d'oubli". C'est le temps au bout duquel les valeurs des observations passées ($\vec{X}_p^* a_p$) n'interviennent plus dans l'estimateur \vec{H}_{k+1} ; il mesure la durée de la mémoire de l'algorithme. On le définira en se donnant un niveau ϵ positif, petit



ESTIMATION LINÉAIRE ADAPTATIVE - APPLICATION AUX TRANSMISSIONS DE DONNÉES

devant 1 tel que

$$(20) \quad \rho_p^k \{u_j\} \leq \varepsilon \quad \mathbb{I}$$

D'après (19), τ_d dépend fortement de p . Ainsi plus les variations de statistique sont tardives, plus l'algorithme à pas décroissants est long à se réadapter.

Considérons maintenant l'algorithme (12)-(14) à pas constants.

On voit alors que

$$(21) \quad \rho_p^k(\mu) = \mu(\mathbb{I} - \mu R)^{k-p}.$$

Or $(\mathbb{I} - \mu R)$ est une matrice définie positive dont toutes les valeurs propres sont plus petites que 1 pourvu que

$$(22) \quad \mu \leq \lambda^{-1}.$$

L'algorithme effectue donc une combinaison linéaire des vecteurs successifs \vec{X}_p^* a_p dont les coefficients décroissent exponentiellement vers zéro en fonction du temps écoulé $(k-p)$. Le temps d'oubli τ_c est uniquement déterminé par μv et ne dépend pas de l'indice p .

Ainsi les algorithmes (12)-(13) et (12)-(14) donnent des estimateurs adaptatifs, capables de poursuivre l'optimalité en présence de variations dans la statistique des signaux. La vitesse des variations qui peuvent être suivies est directement liée à la valeur du temps d'oubli, c'est-à-dire à la vitesse de décroissance des coefficients.

Ainsi l'adaptativité de l'algorithme à pas constants est bien meilleure que celle de l'algorithme à pas décroissants. Nous allons voir dans le paragraphe suivant que c'est au prix d'un manque de finesse dans la convergence.

V - CONVERGENCE DES ALGORITHMES D'APPRENTISSAGE

Comme nous allons le voir, les algorithmes (12)-(13) et (12)-(14) convergent sous les mêmes conditions, mais sur des modes différents. Pour écrire ces conditions, nous définissons la matrice Λ , le vecteur \vec{B} et le nombre non négatif ρ_0 par

$$(23) \quad \begin{aligned} \Lambda &= E \{ (\vec{X}_k^* \vec{X}_k^T)^2 \} \\ \vec{B} &= E \{ \vec{a}_k^* X_k^* \vec{X}_k^T \vec{X}_k^* \} \\ \rho_0 &= E \{ |a_k|^2 \|\vec{X}_k\|^2 \}. \end{aligned}$$

Les conditions de convergence sont

$$(24) \quad \begin{cases} a) \quad \rho_2 \stackrel{\Delta}{=} \text{sup. valeurs propres de } \Lambda < \infty \\ b) \quad \rho_1 \stackrel{\Delta}{=} \|\vec{B}\| < \infty \\ c) \quad \rho_0 < \infty \end{cases}$$

et

$$(25) \quad \begin{cases} \text{les variables aléatoires } (a_k, \vec{X}_k) \text{ et } (a_\ell, \vec{X}_\ell) \\ \text{sont indépendantes pour } k \neq \ell. \end{cases}$$

Notons que cette condition d'indépendance est à coup sûr trop stricte. Bien qu'elle ne soit que rarement vérifiée dans la pratique, on observe la convergence des algorithmes. Mais on n'est pas encore parvenu à s'en affranchir complètement dans les démonstrations théoriques, malgré de récents progrès [7] [8].

V-a) Algorithme à pas décroissants - On peut énoncer le théorème

THEOREME 1 : Si la matrice R est inversible, et sous les conditions (24-25) l'algorithme (12-13) converge presque sûrement, et en moyenne quadratique vers le vecteur complexe \vec{H}_* solution de l'équation de Wiener-Hopf (?).

Pour les signaux et messages réels ce théorème résulte simplement de résultats établis antérieurement, en particulier par Mendel et Fu [3] et par C. Macchi [2]. On montrerait assez facilement que ces résultats se généralisent au cas des fonctions complexes.

V-b) Algorithme à pas constants - On peut énoncer le théorème

THEOREME 2 : Si la matrice R est inversible, et sous les conditions (24)-(25) l'algorithme (12-14) converge en valeur moyenne vers le vecteur \vec{H}_* , pourvu que

$$(26) \quad \mu < \frac{2}{\lambda}$$

En outre la moyenne du carré de la norme $\|\vec{H}_k - \vec{H}_*\|$ tend vers zéro lorsque μ tend vers zéro et k tend vers l'infini :

$$(27) \quad \begin{cases} \varepsilon > 0, \quad \exists \mu(\varepsilon) \text{ tel que} \\ 0 < \mu < \mu(\varepsilon) \quad \limsup_{k \rightarrow \infty} E(\|\vec{H}_k - \vec{H}_*\|^2) < \varepsilon \end{cases}$$

ESTIMATION LINÉAIRE ADAPTATIVE - APPLICATION AUX TRANSMISSIONS DE DONNÉES

On a donc une pseudo-convergence en moyenne quadratique, résultat limite pour μ très petit. Ainsi la convergence de l'algorithme (12-14) est beaucoup moins fine que celle de l'algorithme (12-13). Mais elle est suffisante dans la pratique, ne serait-ce que parce que toute mesure comporte nécessairement une limite de précision. Si l'on désire une très bonne qualité de convergence, on choisira le ϵ de (27) correspondant à cette précision, et donc le μ sera très petit. Mais alors le temps d'oubli deviendra très grand. On voit ainsi qu'on est obligé, dans la pratique, d'adopter un compromis entre les qualités de convergence et d'adaptativité des algorithmes.

DEMONSTRATION : En posant

$$(28) \quad \vec{V}_k = \vec{H}_k - \vec{H}_x$$

il vient d'après (12), (14)

$$(29) \quad V_{k+1} = [I - \mu(\vec{X}_k \vec{X}_k^T)] \vec{V}_k + \mu[-\vec{X}_k \vec{X}_k^T \vec{H}_x + a_k \vec{X}_k]$$

D'après (25) les variables aléatoires (\vec{X}_k, a_k) d'une part et \vec{V}_k de l'autre, sont indépendantes. Prenons la valeur moyenne en tenant compte que \vec{H}_x est solution de l'équation (7). Il vient

$$(30) \quad E(\vec{V}_{k+1}) = (I - \mu R) E(\vec{V}_k)$$

Ainsi

$$(31) \quad E(\vec{V}_k) \rightarrow 0, \text{ si } k \rightarrow \infty$$

si et seulement si la condition (26) est satisfaite (où λ a été défini en (17)). Examinons maintenant

$$(32) \quad \sigma_k^2 = E\{|\vec{H}_k - \vec{H}_x|^2\} = E\{\vec{V}_k^* \vec{V}_k\}$$

En prenant le module au carré de (29), puis sa valeur moyenne, compte tenu de l'indépendance précitée et de l'équation (7) de définition de \vec{H}_x , on trouve, à l'aide des notations (23)

$$(33) \quad \sigma_{k+1}^2 = E\{\vec{V}_k^* \vec{J}(\mu) \vec{V}_k\} + \mu^2 [p_0 + \vec{H}_x^* \vec{A} \vec{H}_x + 2R \{E\{\vec{V}_k^* T(\vec{H}_x - \vec{B}) - \vec{H}_x^* T(\vec{B})\}]$$

Dans cette équation R désigne la partie réelle du nombre complexe, et $\vec{J}(\mu)$ désigne la matrice définie positive

$$(34) \quad \vec{J}(\mu) = I - \mu(2R - \mu A)$$

Grâce aux conditions (24) et à (31), le module du coefficient de μ^2 dans (33) est borné. Soit a la borne. Si $t(\mu)$ est la plus grande des valeurs propres de $T(\mu)$ (33) entraîne donc

$$(35) \quad \sigma_{k+1}^2 \leq t(\mu) \sigma_k^2 + \mu^2 a,$$

et en prenant la limite supérieure sur les indices k

$$(36) \quad (\limsup_{k \rightarrow \infty} \sigma_k^2) \frac{1 - t(\mu)}{\mu} \leq \mu a.$$

Or on sait ([6] p. 63-74) que $t(\mu)$ vérifie

$$(37) \quad t(\mu) = 1 - \mu(2v + \epsilon(\mu))$$

$$(38) \quad \epsilon(\mu) \rightarrow 0 \text{ lorsque } \mu \rightarrow 0.$$

Ainsi il existe des nombres positifs a_1 et μ_1 tels que

$$(39) \quad \forall \mu \leq \mu_1 \quad \limsup_{k \rightarrow \infty} \sigma_k^2 \leq a_1 \mu$$

quantité qui tend vers zéro avec μ . Ceci achève la démonstration du théorème.

VI - APPLICATION AUX TRANSMISSIONS DE DONNEES

VI-a) Le contexte des transmissions de données

En transmission de données, les messages successifs a_k sont des variables aléatoires stationnaires, discrètes, c'est-à-dire prenant seulement un nombre fini de valeurs. Le plus souvent, elles prennent avec égales probabilités des valeurs opposées ce qui rend leur moyenne nulle. Ces messages sont émis à intervalles réguliers, dit intervalles de rythme. Ils sont complexes (cf équation (1)) dans le cas de la modulation de phase ou de la modulation MAQ (on notera que la modulation de phase est un cas particulier de la modulation MAQ). Les deux messages réels en quadrature ont alors même statistique et sont non corrélés si bien que les a_k^2 sont de moyenne nulle. Nous faisons les hypothèses suivantes, en ce qui concerne les données, légèrement plus restrictives que ce que nous venons de dire :

$$(40-a) \quad \left\{ \begin{array}{l} E(a_k) = 0 ; E(a_k^2) = 0 ; E(|a_k|^2) = a^2 \\ \text{les } a_k \text{ sont indépendants et de même loi.} \end{array} \right.$$

Le canal de transmission se comporte comme un filtre linéaire pour le signal transmis. Lorsque la modulation utilisée est linéaire, et c'est le cas de



ESTIMATION LINÉAIRE ADAPTATIVE - APPLICATION AUX TRANSMISSIONS DE DONNÉES

la modulation MAQ, on peut montrer [9] que le signal reçu, après démodulation parfaite à l'aide de deux porteuses en quadrature, est un signal complexe

$$(41) \quad S(t) = \sum_k a_k s(t - k\Delta)$$

où Δ est l'intervalle de rythme et où $s(t)$ est une fonction complexe qui dépend du canal, du modulateur et du démodulateur. En écrivant (41) sous la forme

$$(42) \quad S(k\Delta) = a_k s(0) + \sum_{j \neq k} a_j s(j - k)\Delta,$$

on voit que le signal reçu comporte un bruit d'interférences qui se superpose au signal utile lors de la restitution de la donnée a_k . La fonction $s(t)$, appelée réponse de la ligne est toujours d'énergie finie. Plus précisément nous allons supposer que ses échantillons vérifient

$$(40-b) \quad E \sum_k |s^2(k\Delta)| < \infty.$$

Si en outre, la phase -ou la fréquence- de l'onde porteuse n'est pas parfaitement restituée pour la démodulation, le signal reçu est affecté d'un bruit multiplicatif complexe de module égal à 1, que l'on appelle gigue de phase -et qui peut aussi rendre compte d'une dérive de fréquence-. Enfin l'observation est toujours entachée d'un bruit additif. Ainsi l'observation s'écrit

$$(43) \quad x(t) = (\sum_i a_i s(t - i\Delta)) e^{j\theta(t)} + b(t).$$

Nous faisons les hypothèses statistiques suivantes, sur les messages, le bruit additif et la gigue de phase :

$$(40-c) \quad \left\{ \begin{array}{l} b(t) \text{ est aléatoire, centré, stationnaire, de} \\ \text{puissance finie } \sigma_b^2, \end{array} \right.$$

$$(40-d) \quad e^{j\theta(t)} \text{ est aléatoire, stationnaire}$$

$$(40-e) \quad \left\{ \begin{array}{l} \text{Les éléments } \{ \dots, a_k, \dots \}, b(t), e^{j\theta(t)} \\ \text{sont indépendants} \end{array} \right.$$

$$(40-f) \quad \left\{ \begin{array}{l} \text{Les moments d'ordre 4 des messages et du} \\ \text{bruit sont finis :} \\ b^2 \triangleq E(|a_k|^4) - a^4 < \infty; m^4 \triangleq E(|b(t)|^4) < \infty \end{array} \right.$$

(notons que dans le cas de données, où les a_k prennent un nombre fini de valeurs, la première condition est toujours remplie).

A l'entrée du récepteur l'observation est

échantillonnée au rythme d'émission des données, en vue d'un traitement numérique, puis elle est traitée dans un filtre numérique non récursif autoadaptatif H . Le schéma du système de transmission est porté sur la figure 1.

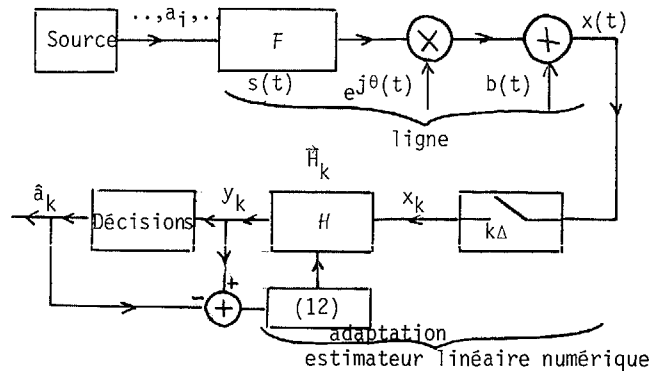


FIGURE 1 : Système de transmission de données

A partir du message estimé y_k , une décision \hat{a}_k est prise sur le message numérique a_k en fonction de la position de y_k dans le plan complexe par rapport aux données possibles.

Lorsque l'estimateur \vec{H}_k est proche de l'optimum \vec{H}_x , l'écart $e_k(\vec{H}_k)$ entre y_k et a_k est le plus souvent très faible, et si le régionnement de \mathbb{C} est correct, la probabilité

$$(44) \quad \text{Pr} [\hat{a}_k = a_k] = 1 - \epsilon$$

de décision correcte est très proche de 1. Dans ce cas on peut continuer de faire marcher les algorithmes adaptatifs après la fin de la période d'apprentissage en remplaçant dans (12) la donnée vraie a_k par la décision \hat{a}_k presque toujours identique. C'est ce qui est illustré sur la Figure 1. On observe alors que malgré les quelques erreurs de décision, les algorithmes (12) continuent de poursuivre l'optimalité de l'estimateur, pendant la phase de transmission de l'information.

VI-b) Etude de l'estimateur optimal

Pour que l'estimateur optimal

$$(45) \quad \vec{H}_x = R^{-1} \vec{T}$$

soit utile, il faut que le vecteur \vec{T} , défini par (9) soit non nul. Or il est immédiat sur (43) que

$$(46) \quad \vec{T} = a^2 E(e^{j\theta(t)}) \vec{S}_0$$

où

ESTIMATION LINÉAIRE ADAPTATIVE - APPLICATION AUX TRANSMISSIONS DE DONNÉES

$$(47) \quad \vec{s}_0^T = (s(-L\Delta), \dots, s(0), \dots, s(N\Delta)) \cdot$$

Ce dernier vecteur est non-nul, en particulier parce que l'instant d'échantillonnage est choisi de telle sorte que $|s(0)|$ soit grand. Quant à la condition

$$(48) \quad G = E(e^{j\theta(t)}) \neq 0$$

elle est en particulier satisfaite lorsque $\theta(t)$ est une fonction gaussienne ou si $\theta(t)$ est la somme de composantes spectrales de phases aléatoires équiréparties et indépendantes, le premier cas correspondant plutôt à la dérive de fréquence entre l'oscillateur local et sa référence ($\theta(t)$ très basse fréquence) et le second correspondant plutôt aux harmoniques du secteur rencontrées dans la gigue de phase.

VI-c) Vérification des conditions de convergence

Nous ne discutons pas ici l'inversibilité de la matrice R_N , que la plupart des auteurs admettent implicitement. Disons seulement qu'en l'absence de bruit additif, les conditions d'inversibilité de R_N sont les mêmes en présence et en l'absence de gigue de phase. Ces conditions sont remplies dans la pratique dans le cas de la modulation MAQ, pourvu que l'instant initial soit choisi convenablement. Enfin, même s'il n'en était pas ainsi, le bruit additif garantirait toujours l'inversibilité de R_N .

En utilisant judicieusement les hypothèses (40-a, ..., f) on peut calculer et borner les grandeurs ρ_0, ρ_1, ρ_2 qui interviennent dans les conditions de convergence (24). On trouve respectivement

$$(49) \quad \rho_0 \leq a^{2(N+L+1)} (a^2 E + \sigma_b^2) + b^2 E < \infty$$

$$(50) \quad \rho_1 \leq |G| a^{2(N+L+1)} \sqrt{E \left(\frac{|b^2 - a^4|}{a^2(N+L+1)} \right) + 2a^2 E + 2\sigma_b^2} < \infty$$

$$(51) \quad \rho_2 \leq |b^2 - a^4| E_2 + (N+L+1)^2 a^2 E^2 \left[2a^2 + 4 \frac{\sigma_b^2}{E} + \frac{m_4}{a^2 E^2} \right] < \infty$$

où G et E_2 sont définis par

$$(52) \quad G = E(e^{j\theta(t)})$$

(valeur moyenne de la gigue)

et par

$$(53) \quad E_2 = \sum_{i=-\infty}^{+\infty} \left[\sum_{j=-L}^N |s^2((i+j)\Delta)| \right]^2$$

quantité qui est finie grâce à l'équation (40-b).

La condition d'indépendance (25) est acquise en espaçant autant qu'il est nécessaire les indices des vecteurs (a_k, \vec{X}_k) utilisés dans l'algorithme.

Ainsi, pour un canal de transmission du type (43) et sous les conditions (40-a, ..., f) concernant les données, la distorsion du canal, le bruit additif et la gigue de phase, les algorithmes (12)-(13), et (12)-(14) convergent au sens des théorèmes 1 et 2 vers l'estimateur linéaire optimal des données.

---:---:---:---:---:---

REFERENCES

- [1] SCHEMETTERER, L. "Stochastic approximation". Fourth Berkeley Symposium, p. 587-609.
- [2] MACCHI, C. "Itération stochastique et traitements numériques adaptatifs". Thèse d'Etat, Paris 1972.
- [3] MENDEL, J.M., FU, K.S. "Adaptive, learning and pattern recognition systems". Academic Press, N.Y., 1970.
- [4] CEA, J. "Optimisation et théorie des algorithmes". Dunod, 1971.
- [5] MACCHI, C., JOUANNAUD, J.P., MACCHI, O. "Récepteurs adaptatifs pour transmission de données à grande vitesse". Ann. Télécom. t. 30, n°9-10, p. 311-330, 1975.
- [6] KATO, T. "Perturbation theory for linear operators". Springer Verlag, New York, 1966.
- [7] FARDEN, D.C. "Stochastic approximation with correlated data". Ph. D. Dissertation, Dpt of Electrical Engineering, Colorado State University, 1975.
- [8] DANIELL, T.P. "Adaptive estimation with mutually correlated training sequences". IEEE Trans. Syst. Sci. Cybernet., SSC-6, 1970.
- [9] MACCHI, C., GUILBERT, J.F. "Réseaux téléinformatiques". Chapitre 3, à paraître, 1977.

